MIT Libraries | DSpace@MIT

MIT Open Access Articles

Experimental assessment of human-robot teaming for multi-step remote manipulation with expert operators

The MIT Faculty has made this article openly available. ***Please share*** how this access benefits you. Your story matters.

Citation: P?rez-D'Arpino, Claudia, Khurshid, Rebecca and Shah, Julie. "Experimental assessment of human-robot teaming for multi-step remote manipulation with expert operators." ACM Transactions on Human-Robot Interaction.

As Published: https://doi.org/10.1145/3618258

Publisher: ACM

Persistent URL: https://hdl.handle.net/1721.1/152999

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of Use: Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.

Massachusetts Institute of Technology

Experimental Assessment of Human-Robot Teaming for Multi-Step Remote Manipulation with Expert Operators

CLAUDIA PÉREZ-D'ARPINO*, Massachusetts Institute of Technology, USA
REBECCA P. KHURSHID†, Massachusetts Institute of Technology, USA
JULIE A. SHAH, Massachusetts Institute of Technology, USA

Remote robot manipulation with human control enables applications where safety and environmental constraints are adverse to humans (e.g. underwater, space robotics and disaster response) or the complexity of the task demands human-level cognition and dexterity (e.g. robotic surgery and manufacturing). These systems typically use direct teleoperation at the motion level, and are usually limited to low-DOF arms and 2D perception. Improving dexterity and situational awareness demands new interaction and planning workflows. We explore the use of human-robot teaming through teleautonomy with assisted planning for remote control of a dual-arm dexterous robot for multi-step manipulation, and conduct a within-subjects experimental assessment (n=12 expert users) to compare it with direct teleoperation with an imitation controller with 2D and 3D perception, as well as teleoperation through a teleautonomy interface. The proposed assisted planning approach achieves task times comparable with direct teleoperation while improving other objective and subjective metrics, including re-grasps, collisions, and TLX workload. Assisted planning in the teleautonomy interface achieves faster task execution, and removes a significant interaction with the operator's expertise level, resulting in a performance equalizer across users. Our study protocol, metrics and models for statistical analysis might also serve as a general benchmarking framework in teleoperation domains. Accompanying video and reference R code: https://people.csail.mit.edu/cdarpino/THRIteleop/

CCS Concepts: • **Computer systems organization → Robotics**; • **Computing methodologies → Robotic planning**; • **Human-centered computing → User studies**; *Interaction paradigms.*

Additional Key Words and Phrases: Teleoperation, Shared Autonomy, Manipulation, Human-Robot Collaboration, Benchmarking

1 INTRODUCTION

Remotely controlling a robot in a distant environment have been a key enabler of robotics in real-world applications where, in addition to navigation and inspection, it is required to interact with the environment in order to change its state. Telemanipulation is particularly relevant in environments where adverse circumstances make it challenging or impossible for humans to be present. Leveraging workflows for human-robot teaming in these situations, by strategically combining human control with the advantages of autonomous systems, have the potential to improve the overall task performance in robotic manipulation, by achieving faster task completion times and higher manipulation dexterity and precision. Examples of important needs for these improvements are

*This work was conducted while authors were at MIT. CPD'A is now at NVIDIA.
†RPK is now at Boston Dynamics.

Authors' addresses: Claudia Pérez-D'Arpino, cdarpino@csail.mit.edu, Massachusetts Institute of Technology, Cambridge, MA, USA; Rebecca P. Khurshid, rkhursh@csail.mit.edu, Massachusetts Institute of Technology, Cambridge, MA, USA; Julie A. Shah, julie_a_shah@csail.mit.edu, Massachusetts Institute of Technology, Cambridge, MA, USA.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
2573-9522/2023/10-ART $15.00
https://doi.org/10.1145/3618258

Just Accepted

Fig. 1. Robot hardware used during the user study. *Left*: Imitation Controller (IC), a passive device used for teleoperation. *Right*: Robot for bimanual manipulation, including two 7-DoF arms, 2 grippers and on-board 2D&3D sensing. The IC is a scaled kinematic replica of the robot, enabling direct teleoperation (**A** & **B**) by commanding the robot to follow the joint angles of the IC on-line. For the teleautonomy conditions (**C** & **D**), the robot is commanded from a computer-based user interface.

found in a wide variety of applications, such as disaster response, explosive ordnance disposal (EOD) [3] [24], space robotics (robotic arm *Canadarm* on the ISS), underwater robotics (underwater manipulation for inspection [19]), and medical applications (surgical robotics [27], teleoperation with *Da Vinci* robot).

The development, evaluation and benchmarking of teleoperation technologies in these challenging domains tend to require *operators with advanced domain expertise* (knowledge particular to the application) and *formal training* (expertise in the use of the interfaces, devices, and methods developed and tested for the domain). These domains are also typically subject to *constrained resources*, such as limited time, *in-situ* deployment, and irreparable damage in case of failure. Altogether, these technical challenges require the development of new technologies that seamlessly integrate human collaboration into the robotic system for augmented perception, planning and control, as well as appropiate benchmarking methods.

In this paper, we present insights from a **user study** and **benchmarking protocol** specifically designed for experimentally assessing the performance of teleoperation methods under realistic field conditions. We perform a within-subjects study with an expert population with ample domain expertise, comparing four telemanipulation frameworks, ranging from direct teleoperation to a human-robot teaming framework with assisted planning. Specifically, we benchmark the following four approaches:

Condition A: Direct teleoperation + 2D perception,
Condition B: Condition A augmented with 3D perception,
Condition C: Teleautonomy interface teleoperation + 2D&3D perception, and
Condition D: Condition C augmented with assisted planning.

The conditions we compare comprise a full software/hardware stack. The robot used is MIT's *Optimus* [29] (Fig. 1(Right)), a dual-arm manipulator (Highly Dexterous Manipulation System (by re^2) with two 3-fingers hands (by Robotiq) and a sensor suite with a Hokuyo sensor (Multisense SL by Carnegie Robotics). The study uses 14-DoF in the arms (7-DoF per arm). The direct teleoperation conditions use an Imitation Controller as input device, depicted in Figure 1 (left). The accompanying video shows the conditions and tasks tested in the study.

We conducted a rigorous statistical analysis of the data collected during the study, and our findings indicate that incorporating a human-robot teaming workflow has significant advantages over direct control. We also hope this benchmarking protocol can serve as a guideline for other researchers conducting detailed evaluations of

Just Accepted

teleoperation and human-robot collaboration models and algorithms, including analysis targeted at understanding the differences between expertise groups. While the statistical analysis in every study must be carefully designed to test the appropriate hypotheses, we make available our R code as a reference for guidance for teleoperation and manipulation studies.

The **results** are consistent with previous findings in the literature [17] of faster execution times for direct teleoperation (**A**) while exhibiting a similar trend in the *trade-off between task time and accuracy* for methods that increase the level of robot assistance and interaction (**B** and **C**). We designed and deployed a human-robot collaboration model based on an assisted planning technique [29] for Condition (**D**), to experimentally assess the advantages of human-robot teaming, which resulted in task times in the range of those produced by direct teleoperation while significantly improving accuracy, overcoming the aforementioned trade-off for the first time without expertly programmed sequences of motions. We also explore the interaction effects between the subject's expertise type and the interface being used and find that while one teleautonomy method exhibits a significant interaction, causing differences in performance per expertise group (**C**), increasing the level of autonomy on the robot's side in a way that furthers the level of collaboration (**D**) overcomes the performance limit correlated with the expertise level, potentially offering a performance equalizer between domain experts with formal training in different fields.

The paper is organized as follows. Section 2 covers related approaches. Section 3 describes our assisted planning approach, and the workflow of each baseline condition is described in Section 4. The protocol details are presented in Section 5, metrics and statistical models in Section 6, the results in Section 7, and the interpretation and analysis in Section 8. Section 9 presents a summary and conclusions.

2 SURVEY OF RELATED APPROACHES

Robotic systems able to operate in **field conditions** for real-world challenging applications are largely based on joint-by-joint teleoperation using switches and joysticks for motion control and 2D camera views for perception [4]. While direct teleoperation enables rapid deployment, this approach has circumscribed deployed systems to the use of robotic arms with a low number of degrees of freedom (DoF), such as the Packbot robot [38]. Increased dexterity based on a higher number of DoF faces scalability problems in terms of what is physically possible to control from this type of interface. The scalability of direct teleoperation is also limited by decreasing performance with the number of joints and instability as time delays increase. Superior levels of dexterity and situational awareness require finding techniques that scale well with the increased workload associated with controlling a higher number of DoF and with managing the information contained in richer perception feedback such as 3D representations of the environment.

The 2012-2015 DARPA Robotics Challenge (DRC) [20] served a large testbed of multiple and competitive approaches to remote robot operation on disaster-response scenarios emulating field conditions. Multiple teams deployed fielded systems [6] [18] [16] to conduct a remote robot through a series of mobility and manipulation tasks inspired by challenges found during the response to the Fukushima nuclear accident in 2011, such as turning a valve or opening a door [25][33]. These instances of telemanipulation systems range in the autonomy spectrum from **teleoperation**, in which the human operator directly controls the movement of the remote robot or of a model of the robot [36], to **teleautonomy**, in which the task is executed through an interaction workflow between the human operator and the robotic system [39][28].

Previous work in the literature analyzes the performance of the teams during the DRC Finals in terms of interaction methods, robot characteristics, control methods, and sensor fusion. Results indicated an increase in performance with increased human robot interaction patterns in terms of balancing tasks between the operator and the robot [28]. While the DRC competition was a state-of-the-art demonstration of the multiple approaches to teleautonomy, and this detailed study found advantages in using human robot teaming strategies, the competition

Just Accepted

conditions made it difficult to conduct a *controlled study*. In particular, teams had different numbers of operators, each with a diverse set of roles within each operational framework. In this paper, we sought to assess the task performance in a controlled fashion with a *single operator* managing all aspects of robot operation.

In **direct teleoperation**, the robot moves simultaneously with the commanded motion using a fixed mapping from the IC to the robot. Multiple human-robot collaboration models that build on the base of direct teleoperation have been proposed and evaluated [32] [2] [34] [31] [13] [17] [5] [30]. In particular, the method of *shared control* aims to improve the usability of teleoperation systems by continuously blending the input signal from the operator with a signal produced by an autonomous system based on a prediction of the objective. This method has been shown to increase task performance in a number of teleoperation settings [5] [23] [14] [22] [15], most commonly in applications where the autonomous assistance is meant to be complementary to operators with some deficit in operating the control interface [23].

Teleautonomy approaches are motivated by the idea of a division of labor between the operator and the robotic system that maximizes their potentials, such as the ability of the robot to perform low-level perception and motion planning, and the operator's high-level task planning and scene understanding. This division of labor is implemented through an *interaction workflow* that enables information to flow between the two parts in order to make planning and execution decisions, typically supervised by the human as the top-level decision maker. The *teleautonomy* workflow is realized through a **teleautonomy interface**, which is a computer interface that affords on-line interactions with a human operator for operations related to perception, planning, and control while immersed in a 3D world that represents the environment of the robot. The most basic planning method on the teleautonomy interface is based on teleoperation, in which the operator can specify the goal pose of the joints or of the end effectors of the robot and the system computes the motion plan to achieve it.

One approach to increase the level of autonomy in the teleautonomy interface is **assisted planning**, in which the robot has the ability to suggest motion plans to the operator automatically in an on-line fashion without the operator indicating the goal explicitly. Taking advantage of the computation capabilities of the robot, it is possible for the system to autonomously compute motion plans that accomplish the task, given that some information about the goal and objects involved can be in the system *a priori* or can be obtained from the operator through on-line queries. A first generation of assisted planning systems in this context deployed during the DRC was based on template-based instances of pre-scripted motion plans executed with supervision from the operators [28]. For tasks known in advance, it is possible to program motions parametrized with respect to objects in the scene and use the human operator in the loop to correctly instantiate these parameters in the scene [21].

A higher level of integration of the human robot team involves overcoming the need for an expert programmer to design sophisticated sequences of parametrized motions in advance. Learning from demonstrations (LfD) is an approach designed to enable robots to learn how to execute manipulation tasks from human demonstrations [1]. Integrating LfD into the system enables a new concept of operations in which a skilled domain expert, not a programmer, can teach manipulation skills to the robot and then execute these tasks remotely in a teleautonomy framework [29]. In previous work, multi-step manipulation tasks have been learned from a single human demonstration using the algorithm C-LEARN [29]. This learning is done by leveraging accumulated knowledge about how humans typically manipulate objects. Tasks are learned in terms of a sequence of steps and a set of geometric constraints that define each step. After the learning phase, the learned task representation is integrated with the teleautonomy framework for human-in-the-loop execution. This strategy results in **teleautonomy with assisted planning**, in which at task execution time, the robot can plan for each learned step and produce a motion suggestion for the human operator. The motions can be generated for new instances of the same learned task where the geometry (position and orientation) of the objects is different from the one in the demonstration.

Previous user studies in the field of surgical robotics have found that teleoperation produces the fastest task times [26] but not the highest accuracy results when compared to models of human-robot collaboration. The user study in [17] compared the use of teleoperation, supervised control, traded control, and full autonomy in an

Just Accepted

inclusion segmentation task. Their results show faster task completion times for teleoperation versus the other methods, while the metric of average force of palpation over the body (less force is desired) was the highest in teleoperation. These previous studies indicate a trade-off between the task execution speed and the accuracy, as measured by task-specific metrics.

This study analyzes similar human-robot collaboration models with various degrees of autonomy, while focusing on the following operational constraints: 1) multiple sequential manipulation steps are required; 2) steps are geometrically constrained; 3) the perception feedback is limited to sensors on board the robot, as opposed to specialized sensor systems mounted externally in a position tailored to the specific task; and 4) tasks are performed by domain experts.

3 ASSISTED PLANNING APPROACH

We seek to benchmark the use of teleautonomy with assisted planning versus other fieldable methods of teleoperation with relevance for the domains of EOD and disaster response. Specifically, we assess the performance of an assisted planning approach based on learned task model using C-LEARN [29], a learning from demonstrations method that enables learning multi-step manipulation skills. In this Section, we first describe the general collaborative workflow and later describe the principles of C-LEARN.

The abstraction of the collaboration workflow consists of combining input from both the robot's planner and the human guidance to create motion plans that can be verified by the user in a 3D animation and approved before robot execution (see Fig.2). This **discrete plan-and-execute workflow** provides enhanced safety for high-risk applications, as the human operator is always aware of the planned movements in advance.

In this workflow (Fig.2), the robot starts in the current keyframe and computes a motion plan to reach the following keyframe in the current scene; this motion plan is displayed to the human operator in the user interface. The operator has the ability to review this plan and approve it for robot execution in a remote environment. Otherwise, the operator can reject the plan and perform modifications by using end effector teleoperation in the interface. Following this logic, the operator can proceed from keyframe to keyframe recommended by the robot system, as illustrated by the green keyframes in Fig.3, or deviate from the robot's plan for a number of keyframes through teleoperation, as represented by the blue keyframes. In the latter case, it is still possible to return to the sequence of suggestions from the learned model, as long as the topology of the task still corresponds to the task learned originally.

For this Condition, we integrated a task representation learned through C-LEARN [29] into this collaborative workflow (Fig.2). A C-LEARN task model is a generalizable representation learned from human demonstrations.

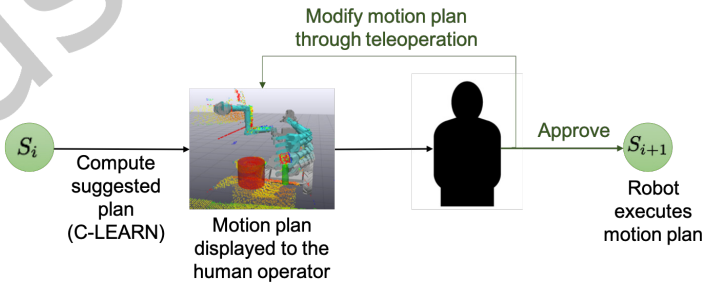


Fig. 2. Human-Robot collaboration model for teleautonomy. The robotic system computes suggested motions plans using a learned task model. Our models were learned using C-LEARN [29], but the collaboration workflow is a general abstraction compatible with any motion generative model that produces sequential discrete muti-step plans.

Just Accepted

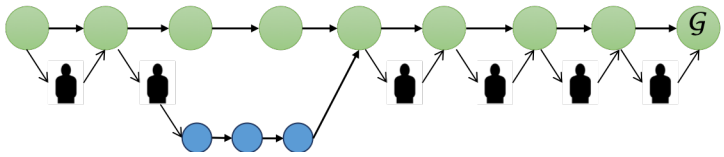


Fig. 3. Workflow of planning and execution for a sequence of keyframes. The operator can accept a suggested motion plan (green keyframes) or perform teleoperation (blue keyframes) and return to suggested plans.

It consists of a series of keyframes, each represented as a set of geometric constraints with respect to objects' frames. These constraints are parameterized, and their specific instances get computed on-line at inference time once the robot receives perceptual input from the environment containing the task-relevant objects (lidar and images). A simplified illustration of a task model is shown in Fig.4. The illustrated task consists of picking up the blue cylinder and dropping it inside the red bucket. A C-LEARN model consists of a set of ordered keyframes (1 to 5 in the example illustration) based on learned parameterized constraints, such as a constrained volume in SE(3) with tolerance, axis orientation constraints (parallel, perpendicular). A C-LEARN keyframe corresponds to a discrete step in Fig.3. Keyframes provide a planner with the necessary information to produce a motion plan given a new topology-preserving scenario consisting in novel positions and orientations of the objects involved in a task with known goal. The learning algorithm has tunable parameters as numerical thresholds related with constraint identification, which can affect the resulting models to be over- or under-constrained [29].

The C-LEARN method first builds a library of motions containing multiple modes for reaching and grasping objects with simple shapes, such as cylinders and boxes. This library is later bootstrapped to acquire a multi-step task model from a single demonstration of a full task. The learned task models used during the user study were learned in advance using data from human demonstrations. These models were learned from seven demonstrations per mode for the library of motions, and one multi-step demonstration per task (Task 1, 2 or 3 in our study).

We selected C-LEARN because (1) the keyframe-based representation made it directly suitable to be integrated into the collaboration workflow, (2) it was designed for tasks with geometric constraints, as is the case for many tasks in our domains of interest, (3) it provides an interpretable and verifiable task model. These were design choices made in our previous work [29]. Alternatively, future developments of other learning techniques that use a discrete task model in the form of keyframe could be integrated in a similar manner.

While a system with this task model could theoretically be deployed autonomously, in practice, a high task success rate requires a human operator in the loop. This need is due to the accumulated errors from on-line perception (pose estimation, occlusions, etc), task-level planning capabilities, partial observability (only on-board

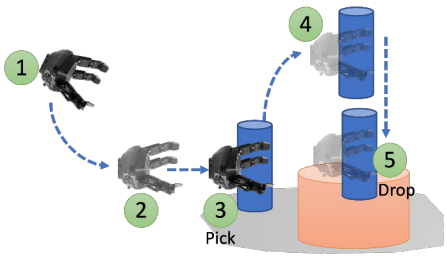


Fig. 4. Illustration of a keyframe-based task sequence. The representation based on keyframes results in a parameterized flexible task model, whose sequence of steps can be instantiated in new scenes as a function of object poses.

Just Accepted

sensors) and the need to respond to uncertain outcomes in the real world that are better discerned by the human operator.

Note that this paper evaluates the performance of this interaction workflow that uses a learned task model, but not the capabilities of C-LEARN as a learning model. It is challenging to perfectly disentangle the effects of the collaboration workflow itself versus the contributions of the underlying task model. On one hand, the collaboration workflow doesn't depend on how the underlying task model was built (inner-workings of the learning algorithm), as long as the learned representation can be fit into the workflow. On the other hand, the quality of the underlying task model directly impacts the performance of the collaboration workflow, as providing feasible and correct motion suggestions is one of the principles that enables improving task performance, and the frequency of usable suggestions increases for better task models. Taking this into account, we believe the insights regarding the performance improvements offered by assisted planning vs teleoperation would remain even if changing the underlying learning model, as long as the model is at least of similar quality (i.e. equivalent likelihood of generating feasible and correct motion plans) as the models we tested.

Just Accepted

4 CONDITIONS

In order to experimentally assess the contributions of assisted planning with C-LEARN, we conduct the user study covering a range of methods representative of fieldable teleoperation techniques with increasing levels of autonomy and enhanced perception. The set up for each condition is summarized in Figure 5. Examples of the interface workflow, robot task execution and OCU room are shown in the accompanying video for all conditions.

Conditions **A** and **B** (Figure 5) are based on **direct teleoperation** using an *imitation controller (IC)* device (Figure 1 Left). The IC is a passive device whose structure is a scaled kinematic replica of the robot, which enables motion of the IC to be directly mapped to robot motion. We explore the use of the IC with two variants of perception: (**A**) 2D perception and (**B**) 2D+3D perception, in which 3D perception consists of a view as described in Figure 5. **2D Perception** provides the following four 2D camera live views: one camera mounted on each wrist of the robot, one on the base, and one on the head.

Conditions **C** and **D** are based on the **teleautonomy interface**, depicted in the screen views in Figure 5(c)(d). We use the interface *Director* [21], an open-source user interface developed to pilot the Atlas robot in the DARPA Robotics Challenge (DRC) Finals. Through this interface, the operator has access to a 3D representation of the robot model and the robot's environment (Figure 5). This 3D representation is displayed in a 2D monitor, similarly to the 3D environment used in video games or other robot interfaces such as RVIZ from the Robot Operating System (ROS). In the teleautonomy interface, we experiment with the following two planning workflows. Condition **C** is based in end-effector teleoperation, in which the operator indicates the desired pose of the end effectors of the robot by either manually dragging the virtual robot's hands or by positioning virtual floating hands within the 3D view in the interface. Condition **D** uses assisted planning based on C-LEARN[29], in which the system displays to the operator a series of suggestions of motion plans automatically.

During motion execution in all methods, the low-level controller of the robot used position control in joint angles space. The robot arms were connected to constant power in all runs to avoid the possibility of performance differences due to decreasing battery levels during the course of the study.

4.1 (Condition A) IC-based direct teleoperation + 2D perception:

During *IC-based direct teleoperation*, the operator manipulates the Imitation Controller device (IC) (Figure 1). Joint angles from the IC are mapped directly to joint angles in the real robot, which executes the commanded motion simultaneously with the IC. **2D Perception** provides the following four 2D camera live views: one camera mounted on each wrist of the robot, one on the base, and one on the head. Joint angles are streamed from the IC

Fig. 5. View of the Operator Control Unit (OCU) room (left column), and the content displayed on the computer monitor for each condition (right column):
Condition A: IC-based direct teleoperation + 2D perception;
Condition B: Condition **A** augmented with 3D perception;
Condition C: Teleautonomy interface teleoperation + 2D&3D perception;
Condition D: Condition **C** augmented with assisted planning.
The operator had no direct line of sight with the robot in any condition, and the situational awareness of the operator is restricted to the perception feedback provided through the system. 2D perception is provided by four camera feeds located on the top of the screen, and 3D perception is provided by the rendering of a 3D robot model, 3D point clouds, and rendered 3D virtual objects from pose estimation. The user interface is based on *Director* [21].

To Be Accepted

to the robot at 10 Hz, and camera feed is streamed from the robot to the OCU screen at 15 frames per second (fps). A view of the interface is shown in Figure 5(a).

This condition provides a field benchmark, as it uses similar perception as in the systems fielded today for real operations. In terms of the input device for teleoperation, field operations use switches and joysticks to control individual joints, but this methodology is used with three or four degrees of freedom (DoF) only. Similar to the Imitation Controller, a Joystick device allows either position or velocity-based end effector control, but we acknowledge that this method does not scale well with number of DoF, making the comparison unfair when using a high DOF robot. We choose an IC as it provides an augmented sense of the position of the entire arms, enabling the user to directly perform modifications per joint if desired (this could be done with a joystick but would require mode change per joint). The IC allows human teleop on both end effector and joint space simultaneously, an advantage that can be useful for high DoF arms, by enabling control of all joints or simply moving the end effector of the IC and leaving all the IC joints accommodate accordingly.

4.2 (Condition B) Condition A augmented with 3D perception:

This condition consists of the same IC-based teleoperation implementation as in condition **A**, with the addition of 3D perception. In addition to the live 2D views available in Condition **A**, **3D perception** provides a live view of the 3D point cloud being sensed by a rotating Hokuyo mounted on the head of the robot, and a visualization of the robot model from sensed joint angles.

4.3 (Condition C) Teleautonomy interface teleoperation + 2D&3D perception

The teleautonomy interface enables the operator to command the robot through **end effector teleoperation**. Unlike IC-based direct teleoperation, where the robot moves simultaneously with the motion of the IC input device, teleautonomy has a two-step workflow of planning and execution. Motion plans are first composed on the interface and are executed on the real robot only after approval. The workflow for this condition is shown in Figure 6.

The current state of the robot and the environment is represented on the 3D view. The operator specifies a desired goal position and orientation of the end effectors of the robot. The goal can be specified by either dragging the hands of the robot on the interface or by locating a floating hand. After the goal is specified, the system automatically computes a motion plan for moving from its current configuration to a configuration that satisfies the requested end effector pose. The system produces a 3D animation of the computed motion plan, which is shown overlapped with the 3D representation of the environment. The animation can be re-played by the operator as many times as needed. The operator approves or rejects the motion plan. If approved, the motion plan is sent to the robot for execution.

4.4 (Condition D) Condition C augmented with assisted planning:

This condition focuses on **assisted planning**, in which the robot automatically suggests a sequence of motion plans to the operator. The workflow is illustrated in Figure 7. Given a start state, the operator clicks a button to obtain the next motion suggestion. The suggestion is previewed in animation, in the same fashion as in condition **C**. If the operator approves, the motion plan is sent to the robot for execution. This workflow illustrated in Figure 7. Additionally, the operator is free to recur to the end effector teleoperation workflow (**C**) at any moment.

In this study, the motion recommendations are generated from a pre-learned model using C-LEARN [29] (See Section 3). For the sake of uniformity, the pre-learned task models are the same for all participants, ruling out performance differences from the models themselves. While we reuse the same models, note that motion suggestions are generated on-line by the model according to the current environment. Participants had not seen in advance any motions that the robot would suggest for each task.

Just Accepted

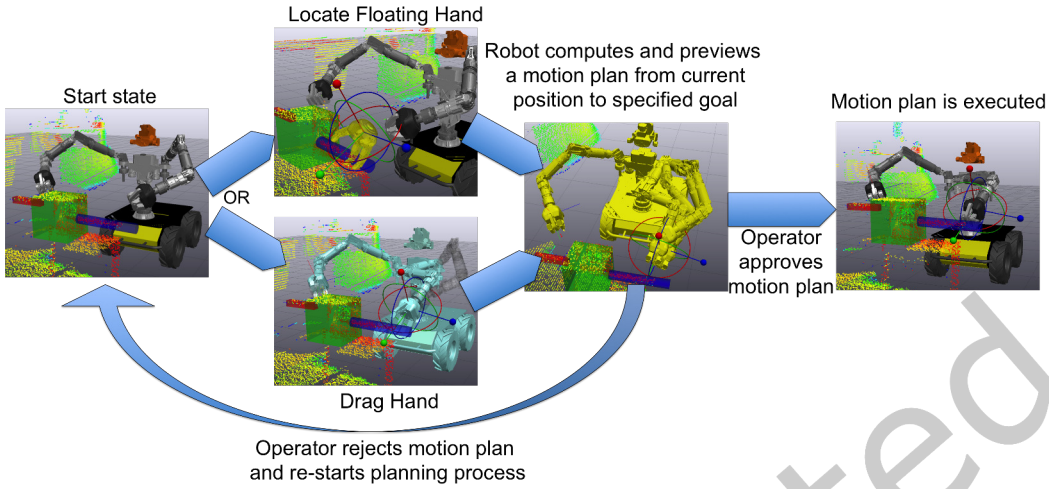


Fig. 6. Workflow for end effector teleoperation in the teleautonomy interface. Available in conditions **C** and **D**.

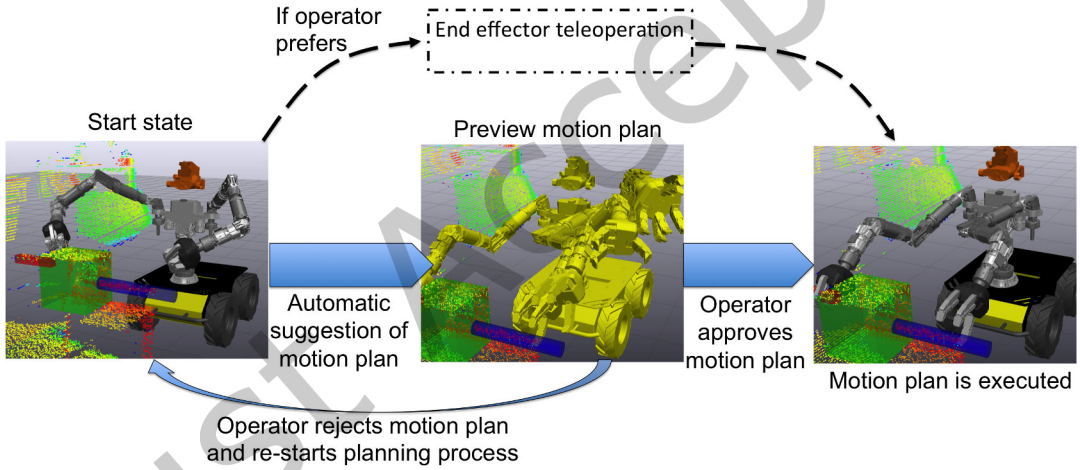


Fig. 7. Workflow for assisted planning. Condition **D**.

Unlike IC-based teleoperation (conditions **A** and **B**), where there is simultaneous motion as the robot tracks the IC commands, in the teleautonomy interface (conditions **C** and **D**), motion plans are first elaborated and previewed on the interface and then sent to the robot for execution upon approval from the operator. Both **C** and **D** use the same optimization-based motion planner [6] [29], available in Drake [35], which uses the solver SNOPT [7]. The difference depends on the source of the goal specification (operator vs. automatic).

Note that *Director* is used for 3D perception on Conditions **B**, **C** and **D**. In **B**, Director affords only visualization, while in **C** and **D** it affords visualization and interaction for robot control. In Condition **C**, the operator uses the visualization of the sensor data (point cloud and virtual objects) as a reference to specify the desired position of the end effectors. In assisted planning in **D**, the robot computes automatically the goal of the end effectors with

Just Accepted

respect to frames in the objects, which are obtained from the same perception subsystem. The perception-based pose estimation of objects is the same in both conditions.

5 USER STUDY DESIGN AND PROTOCOL

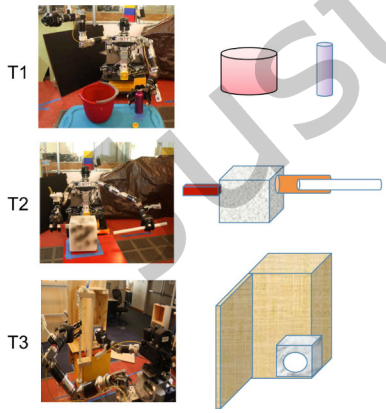
A within-subjects study with an expert population in robot telemanipulation was conducted to evaluate the performance of four interfaces in three manipulation tasks. We collected objective performance data during the task executions and subjective performance and satisfaction metrics with a number of surveys. The study was conducted over the course of two days per participant. This time permitted extensive training per condition and frequent breaks. All participants executed three *Tasks* (T1, T2 and T3) in all four *Conditions* (A,B,C,D) during two successful *Trials* of each task, for a total of 24 runs per participant, and 288 runs in the complete user study. Each trial used the same initial pose of the robot and the same start position and orientation of the objects involved in each task. Each participant followed the protocol outlined in Figure 8(middle). The study protocol and consent form were approved by the Institutional Review Board of the Massachusetts Institute of Technology.

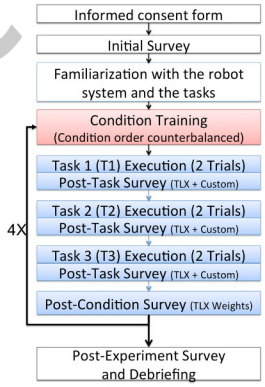
5.1 Participants and Assignment Method

The study was conducted with a total of 12 participants (11 males, 1 female, aged 24-41, M=30.83, SD=4.82), recruited from an **expert population with domain expertise and practical experience with remote robot control**. A summary of the self-reported expertise collected on the initial survey is presented in Figure 8(right). The expertise level of the participants is divided in the following two groups:

ONR: 6 participants (5 males, 1 female, aged 30-41, M=34.17, SD=4.36) with domain expert knowledge in the area of EOD. The ONR group was recruited in collaboration with the **O**ffice of **N**aval **R**esearch (ONR). In particular, 3 of the participants in this group are professional EOD technicians. This group has extensive expertise in joint-by-joint teleoperation (primarily using switches and joysticks) of low-DOF robots (e.g., iRobot Packbot [38], Foster-Miller Talon) using 2D perception only (from on-board cameras). This group is professionally trained to execute complex telemanipulation tasks under time pressure and safety concerns as required in EOD.

DRC: 6 participants (6 males, aged 24-30, M=27.5, SD=2.35) with domain expertise in Robotics. The DRC group was recruited from operators that participated in the DARPA Robotics Challenge (DRC). This





ONR Group: Professional EOD technicians

DRC Group: Robot pilots for the DARPA Robotics Challenge

	ONR (M; STD)	DRC (M; STD)
Age [years]	34.17; 4.36	27.5;2.35
Expertise in robot teleoperation [1-5 scale]	3.83; 0.98	3.83; 0.98
Hours logged remotely controlling robots [Hours]	**245; 179.08** (40,80,250,200,400,500)	**204; 173.71** (24,100,100,200,300,500)
Used Director to control remote robots [Hours]	0.00; 0.00	113; 96.53
Used Imitation Controller to control remote robots [Hours]	10.33; 16.50	0.08; 0.20
Familiar with Director [1-5 scale]	1.33; 0.82	3.83; 0.98
Familiar with Imitation Controller [1-5 scale]	2.83; 1.33	1.33; 0.52
Experience playing video games and other 3D virtual environments [1-5 scale]	3.50; 0.55	4.17; 0.75

Fig. 8. Manipulation tasks (left). Study protocol flow diagram (middle). Participants' information as self-reported in the initial survey (right).

Just Accepted

group has experience in remote control of high-DOF robots (e.g., Atlas robot by Boston Dynamics), with 3D perception through the user interface *Director* [21]. This group is trained to execute telemanipulation tasks under time pressure as required for the DRC competition.

While both expertise groups have an average of over 200 hours of experience in remotely controlling robots, they are different in terms of the types of interface they most commonly use. The ONR group is more experienced with direct teleoperation devices, whereas the DRC group is more familiar with computer-interface based teleoperation.

The order in which participants of each expertise group (DRC and ONR) experienced the four conditions was selected using a 4x4 Latin square. For each group, 4 participants completed one 4x4 Latin square, and the 2 remaining were assigned a random row. The Latin squares of each expertise group were counter-balanced. A manipulation check over task time showed no significance for the effect of the order in which conditions were experienced ($p = 0.34$).

5.2 Tasks

In each condition (randomized), three tasks (Figure 8(left)) were executed in the same sequential order (not randomized). We designed the tasks to be ordered in increasing order of difficulty, according to previous experimentation, and to cover a variety of important manipulation skills: grasping, transport, precise positioning of hands, constrained motion, bimanual manipulation, and reaching into confined spaces.

Task 1 (T1): Grasp the cylinder at the left, transport it, and release the cylinder inside a container. This task requires only one arm.

Task 2 (T2): Grasp the handle at the right side to secure the box, grasp the cylinder at the left, and extract the cylinder. This task requires dual arm manipulation with simultaneous contact with the same structure and satisfaction of constraints in motion (cylinder extraction).

Task 3 (T3): Grasp the door handle (rope) at the left, open the cabinet (articulated object), release the door handle, reach inside the cabinet at the right side, and push a button to turn on a light. This task requires dual arm manipulation and reaching into a confined space.

5.3 Training

Participants were trained for each condition immediately before the test time in that condition. No time limit was assigned for training sessions to enable participants to achieve the expected level of proficiency before proceeding with the test tasks. All participants had a guided training session, during which the same instructor provided the same sequence of technical instructions interleaved with practice time for each concept or technique involved. Following the same sequence of instructions with all participants resulted in consistent training interaction and timing across participants. After the guided training session, participants practiced using all capabilities in the interface to execute a training task. Participants were allowed to proceed to the test session after being able to execute the training task successfully without technical errors in the use of the interface and having expressed feeling comfortable with it. The training task consisted of grasping, transporting, and releasing a cylinder over a table. The task was practiced with both arms, and the objects were located so that motions had to take place in a region of the workspace similar to the one used in the test tasks. During training, participants had available live video feedback of the robot room from cameras external to the robot with the purpose of facilitating training.

6 METRICS AND STATISTICAL MODELS

We investigate and compare the performance of Conditions **A**, **B**, **C** and **D** through a number of objective and subjective metrics. The metrics and statistical models used for analysis are summarized in Table 1 and described below.

Just Accepted

Table 1. Metrics and statistical models

Metric	Definition	Model
Total Task Time	Total task time spans the time from the moment the operator takes control of the interface to the moment the task is accomplished	GLME with Gamma distribution (logarithmic link) $DV \sim Trial + (Condition * Expertise) + (Condition * Task) + (Task * Expertise) + (1\|Subject\_ID)$ (1)
Object of Interest (OOI) Moved	Number of times (count) the robot moved the OOI (manipulation target) outside the task instructions. For example, due to undesired displacements during pushing or lifting.	GLME with Poisson distribution $DV \sim Trial + Condition + Expertise + Task + (1\|Subject\_ID)$ (2)
Collisions with Other Objects	Number of times the robot came into unintended contact with objects other than the OOI.	
Collisions with OOI	Number of times the robot came into unintended contact with the object of interest.	
Re-grasps	Number of times the operator reattempted the same grasp. For example, the operator commanded the hand to close with the intention of grasping an object, but the hand closed in free space due to incorrect positioning of the end effector and the grasp had to be re-attempted.	GLME with Poisson distribution $DV \sim Trial + Condition + Expertise + (1\|Subject\_ID)$ (3)
Full Grasps Vs. Tip Grasps	Every grasp was classified into two categories. Full grasps are defined as stable grasps where the object is in contact with the palm of the hand and all three fingers achieved closure. A tip grasp is defined as a grasp that held the object but failed to satisfy the full grasps conditions.	GLME with binomial distribution (logarithmic link) $DV \sim Trial + Condition + Expertise + Task + (1\|Subject\_ID)$ (4)
NASA Task Load Index (TLX)	Mental Demand, Physical Demand, Temporal Demand, Effort and TLX Total Score as defined by the NASA **T**ask **L**oad Inde**X** [8], resulting in seven scores in the continuous range 0-100 (bounded).	GLME with binomial distribution $DV \sim (Condition * Expertise) + (Condition * Task) + (Task * Expertise) + (1\|Subject\_ID) + (1\|Row\_ID)$ (5)
HRI Metrics in Post-Condition Survey	Likert Scale questions are grouped in six categories (see Table 2): Robot teammate traits[9], Working Alliance – Bond Subscale[12] [9], Working Alliance – Goal Subscale [12] [9], as well as custom metrics for Manipulation, Perception, and Satisfaction. The response range is 1-7 (discrete), and the model assumes the observed discrete numbers are a proxy for the underlying continuous (and unbounded) scale.	Linear Mixed Effect Models $DV \sim (Condition * Expertise) + (Condition * Task) + (Task * Expertise) + (1\|Subject\_ID)$ (6)
Condition Ranking in Post-Experiment Survey	After completing the study, participants were asked to rank all conditions according to their preferences regarding nine performance-related aspects and one overall final ranking	Ordinal cumulative link mixed model $DV \sim (Condition * Expertise) + (1\|Subject\_ID)$ (7)

Just Accepted

6.1 Independent variables (IV)

Condition (interfaces **A**, **B**, **C** and **D** as categorical variable); ***Task*** (Tasks T1, T2 and T3 as categorical variable); ***Expertise*** (Expertise levels DRC and ONR as categorical variable); ***Trial*** (Trial number Trial 1 and Trial 2 as categorical variable); ***SubjectID*** (Subject identifier 1 to 12 as categorical variable).

6.2 Dependent variables (DV)

Objective metrics: **Total Task Time**, **OOI moved**, **Collision with OOI**, **Collision with other objects**, **Full Grasps Vs. Tip Grasps**, **Regraps**).

Subjective metrics: seven **TLX scores**, seven **HRI scores**, ten final **Condition Rankings**).

See Table 1 for the definition of each metric.

6.3 Models

The effect of the IVs considered in this study on each DV is determined by fitting a **mixed effect model** individually for each DV. Each model included random effects of the factor **SubjectID** with a random intercept $(1|SubjectID)$, to account for different responses per subject according to their baseline level. Each model tested the fixed effects of *Condition*, *Task*, *Expertise*, *Trial* for main effects, simple effects, and 2-way interactions when pertinent, as detailed below. Holm adjustment was applied to the models [11]. Table 1 describes each model [1] and equation using the Wilkinson notation [37].

For **Total Task Time**, we hypothesized the following effects: main effect of *Trial* to account for the learning effect of using the same interface in the same task during the two trials; 2-way interaction for $(Condition * Expertise)$ to account for time performance differences in the same condition between participants from the ONR and DRC groups; interaction for $(Condition * Task)$ to account for a given condition enabling a level of dexterity in a specific manipulation skill present in one task but not others that resulted in time performance differences; $(Task * Expertise)$ to account for possible previous expertise causing a time performance difference across different tasks. Three-way interactions were not hypothesized, and no significance was found when tested. This model is represented in Eq.1 in Table 1.

The data of Total Task Time exhibited strong heteroscedasticity. We used a generalized linear mixed effect model with a gamma distribution (logarithmic link) to account for this heteroscedasticity.

For all other objective metrics – **OOI moved**, **Collision with other objects**, **Collision with OOI** (Eq.2 in Table 1), **Re-graps** (Eq.3 in Table 1), and **Full/Tip Grasps** (Eq.4 in Table 1) – we hypothesized the main effects of *Trial* for potential learning effects, *Task* to account for different geometries changing the likelihood of different events, *Condition* to account for the interface, and *Expertise* to account for performance differences due to pre-existent skills. For the **Re-grasps** metric, the factor *Task* was removed due to non-convergence (Eq.3).

The metrics **OOI moved**, **Collision with other objects**, **Collision with OOI** (Eq.2), and **Re-graps** (Eq.3) were fitted individually using a generalized linear mixed effect (GLME) model with a Poisson distribution to model counts. The metric **Full/Tip Grasp** was modeled with a generalized linear effect model with a binomial distribution (logarithmic link), which enables use of the counts of Full and Tip Grasps to model the probability of a Full grasp (Eq.4).

Since the objective metrics other than Time consider discrete events that occur sparsely during a task run, it's unlikely to observe significance for interaction terms from this data. Thus, two-way interactions were not hypothesized and tested not significant when checked.

For the **NASA Task Load Index (TLX)** [8], we hypothesize the existence of two-way interactions for the factors *Condition*, *Expertise* and *Task*. Note that *Trial* is not a factor because the TLX questionnaire was administered only after participants finished the second trial of each task. The model results in the equation Eq.5 in Table 1. We fit a generalized linear mixed model with a binomial distribution to each of the seven scores (range 0-100).

For the **HRI Metrics in Post-Condition Survey** (see questionnaire on Table 2), we fit a linear mixed model to each group with the equation Eq.6 in Table 1. We also report the Cronbach's α measure of consistency for each group.

[1]Statistical support was provided by the Institute for Quantitative Social Science, Harvard University.

Just Accepted

For the **Condition Ranking** in the in Post-Experiment Survey, we fit an ordinal cumulative link mixed model was fit to each ranking question. We hypothesized a 2-way interaction of *Condition* and *Expertise*. Note that *Task* and *Trial* are not factors for this model because this information was queried by the end of the user study. The model uses the equation Eq.7 in Table 1.

Data analysis was performed using R (R version 3.4.2) using custom code and R packages lsmeans_2.25, ordinal_2015.6-28, ggplot2_2.2.1, effects_4.0-0, optimx_2013.8.7, robustlmm_2.1-3, nlme_3.1-131, lattice_0.20-35, lmerTest_2.0-33, car_2.1-5, lme4_1.1-14. No treatment for outliers was performed.

7 RESULTS OF THE STATISTICAL ANALYSIS

In this Section we present the key insights into the numerical results from the statistical analysis for all objective and subjective metrics. Each result is identified as R_i to facilitate cross reference in the analysis presented in Section 8.

Results for each metric are accompanied by a summary Figure that presents (1) a bar plot for the metric response per condition (expected values listed numerically) with error bars (Standard Error); (2) tables with the p-values from condition contrasts (e.g. $Condition_i$ vs $Condition_j$ for the same factor level and $Condition_i$ vs $Condition_i$ across different levels), and (3) analysis of deviance table, indicating which model factors were significant for main effects and interactions. Each plot shows an arrow indicating the ideal directionality of the y-axis variable.

7.1 Objective Metrics

Total Task Time (See Fig.9)

How do Conditions A, B, C and D compare?

Table 2. HRI questions in post-condition survey

Robot Teammate Traits:
"The system was intelligent"
"The system was trustworthy"
"The system was committed to the task"
Working Alliance – bond subscale:
"I felt physically uncomfortable using the system" (reverse)
"The system and I understand each other"
"The system and I respect each other"
Working Alliance – goal subscale:
"The system perceives accurately what my goals are"
"The system does not understand what I am trying to accomplish" (reverse)
"The system and I are working towards mutually agreed upon goals"
"I find the system's actions confusing" (reverse)
Manipulation:
"How confident were you in your ability to move the robot's arm to a desired location?"
"How well did the robot's motion match your intended motion?"
"I was easily able to manipulate objects that were in close proximity to other objects"
"I was easily able to control the robot and objects held in free space"
"I was able to move the robot's arm accurately and precisely"
"The robot did what I wanted"
"I was able to accomplish the task quickly"
Satisfaction:
"I was satisfied by my performance"
"I would use this system the next time the task were to be completed"
"For this task, I would recommend to use this condition in the field"
"How would you rate your overall experience"

Just Accepted

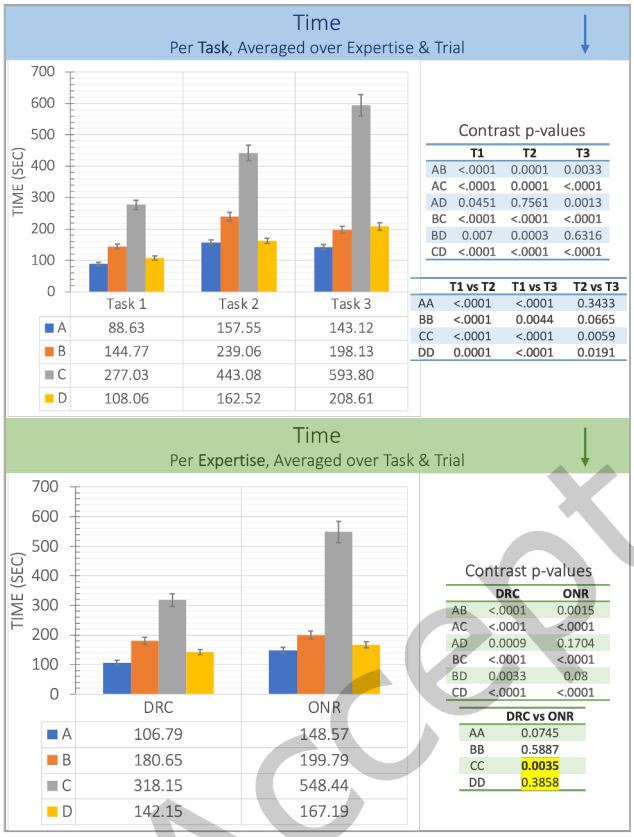


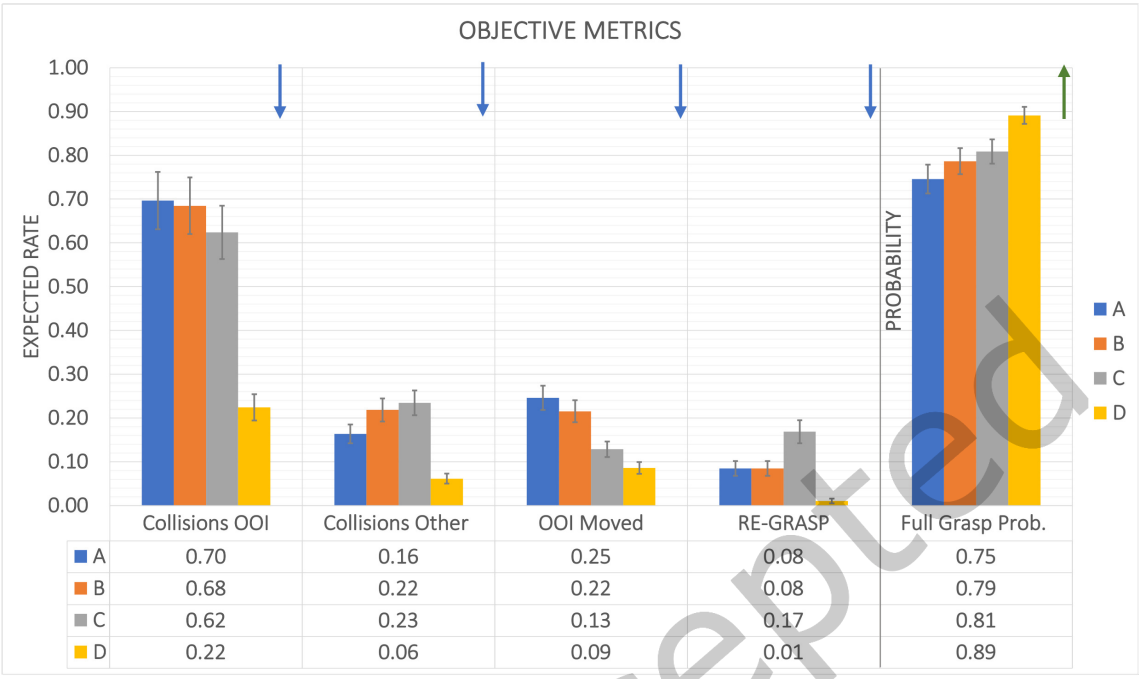
Fig. 9. Total Task Time. Top chart: **Time per Condition per Task** averaged over all Expertise and Trial levels. Bottom chart: **Time per Condition per Expertise** levels averaged over all Task and Trial levels. Error bars show the Standard Error (SE). The tables summarize the p-values from condition contrasts next to the corresponding plot.
The Analysis of Deviance Table shows the significance of the factors in the model. The arrows show the ideal directional of the y-axis variable. For example, faster task times are better.

R1: D was significantly faster than **C** for all task ($p < 0.0001$) and expertise ($p < 0.0001$) groups.

R2: D total time is within a range comparable to teleoperation **A** and **B** with no significant difference (Fig.9). The expected means follow the relation $A < D < B$ for all tasks and expertise levels except for T3 where $D > B$ with no significant difference ($p = 0.6316$).

R3: C took more time than **A**,**B**, and **D** for all tasks and expertise groups ($p < 0.0001$).

Just Accepted



	Collisions OOI	Collisions Other	OOI Moved	RE-GRASP	Full Grasp Prob.
A	0.70	0.16	0.25	0.08	0.75
B	0.68	0.22	0.22	0.08	0.79
C	0.62	0.23	0.13	0.17	0.81
D	0.22	0.06	0.09	0.01	0.89

Contrast p-values

	Collisions OOI	Collisions Other	OOI moved	Re-Grasp	Full Grasp
AB	1.0000	0.5718	0.5640	1.0000	0.9912
AC	1.0000	0.5324	0.0789	0.3440	0.8424
AD	0.0001	0.0478	0.0043	0.2426	0.0371
BC	1.0000	0.7719	0.2103	0.3440	0.9912
BD	0.0001	0.0038	0.0188	0.2426	0.1841
CD	0.0006	0.0021	0.5126	0.0423	0.3471

Analysis of Deviance Table (Type III Wald chisquare tests)

Pr(>Chisq)	Collisions OOI	Collisions Other	OOI Moved	Re-grasp	Full Grasp Prob.
(Intercept)	0.02854 *	5.40E-09 ***	8.93E-09 ***	2.96E-08 ***	0.111767
Trial	0.08888 .	0.19191	0.563903	0.477918	0.208448
Task	7.25E-05 ***	5.90E-16 ***	<2.20E-16 ***	N/A	0.002212 **
Condition	9.78E-05 ***	0.002727 **	0.002034 **	0.024559 *	0.05063 .
Expertise	0.30306	0.524511	0.008907 **	0.006104 **	0.107252

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Fig. 10. Objective metrics. Error bars show the Standard Error (SE). The tables summarize the p-values from condition contrasts for each metric. The Analysis of Deviance Table shows the significance of the factors in the model.

R4: Expected mean time of **A** is lower than **B** for all task ($T1, T2 : p < 0.0001; T3 : p = 0.0033$) and expertise levels ($DRC : p < 0.0001, ONR : p = 0.0015$).

How does Condition interact with Expertise and Task?

R5: The results show a significant two-way interaction for **Condition*Expertise** ($Pr(> Chisq) = 0.0009$). The bottom chart in Fig.9 shows the Task Time per Condition per Expertise levels, averaged over all Task and Trial levels. **C** is the only condition that experimented a significant Task Time 2-way interaction between the ONR and DRC expertise groups, taking and expected mean increase of 1.72 for the ONR group ($p = 0.0035$) (Fig.9 bottom). This performance difference might be attributable to the previous experience in controlling robots through a computer interface by the DRC group.

R6: However, while using the same computer interface as in **C**, condition **D** resulted in no significant time difference between the two expertise groups ($p = 0.3858$) (Fig.9 bottom), providing evidence that

Just Accepted

Table 3. Significance of the model factors for the TLX Metrics

Analysis of Deviance Table (Type III Wald chisquare tests)

Pr(>Chisq)	TLX Total Score	Mental Demand	Physical Demand	Temporal Demand	Performance	Effort	Frustration
(Intercept)	0.123890	0.9584456	0.9532	0.31456	0.005533 **	0.236965	0.018727 *
Condition	2.28E-09 ***	0.0005099 ***	<2e-16 ***	0.01948 *	0.084293 .	6.56E-08 ***	0.009681 **
Task	0.930809	0.3593196	0.4940	0.95917	0.139784	0.689194	0.424621
Expertise	0.024180 *	0.0008778 ***	0.2801	0.02925 *	0.002198 **	0.009646 **	0.007391 **
Condition:Task	0.524299	0.2786167	0.7920	0.74798	0.010695 *	0.387657	0.021860 *
Condition:Expertise	0.003931 **	0.0169668 *	0.6737	0.05831 .	0.114086	0.002364 **	0.015707 *
Task:Expertise	0.933050	0.9976282	0.9305	0.74269	0.622153	0.935531	0.991234

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

assisted planning makes complex interfaces more accessible to different training backgrounds. Similarly, for direct teleoperation conditions **A** and **B**, Expertise resulted in no significant differences in performance ($A : p = 0.0745; B : p = 0.5887$).

R7: The results show significant two-way interactions for the pairs **Condition*Task** ($Pr(> Chisq) = 0.0046$). The top chart in Fig.9 shows the Task Time per Condition per Task, averaged over all Expertise and Trial levels. For interface conditions (C,D), total time followed $T1 < T2(C, D : p < 0.0001)$, and $T2 < T3(C : p = 0.0059; D : p = 0.0191)$, while direct teleoperation conditions (A,B) resulted in $T1 < T3(A : p < 0.0001; B : p = 0.0044)$, and $T3 < T2(A : p = 0.3433; B : p = 0.0665)$.

Is there a learning effect between trials of the same task?

R8: The second trial resulted in faster times than the first trial when averaged across all expertise and task levels $p < 0.0001$. We attribute this result to a learning effect on how to execute a particular task using a particular interface. However, a manipulation check with respect to total task time showed that this learning effect was not significant across conditions ($p = 0.34$), meaning the learning effect of task X with interface Y did not carry over to task X with interface Z.

OOI-Moved, Collisions OOI, Collisions Other objects & Grasping (See Figure 10)

R9: In addition to Task Time, the rest of objective metrics showed benefits of using Condition **D** over the other method. The expected rate of **OOI-Moved** decreased as $A > B > C > D$. The rate for **D** is significantly lower than for both of the direct teleoperation ($A : p = 0.0043; B : p = 0.0188$). The rate of **Collision-OOI** for **D** is significantly lower than all conditions ($A : p = 0.0478; B : p = 0.0038; C : p = 0.0021$). The rate of **Collisions-Others** for interface condition **D** is significantly lower than both of the direct teleoperation conditions and the other interface condition C ($A : p = 0.0478; B : p = 0.0038; C : p = 0.0021$). For the grasping metrics, the expected **re-grasp** rate difference between **A** and **B** was not detectable (p=1.000), whereas for interface conditions **D** had a smaller rate than **C** ($p = 0.0423$). The expected probability of a full grasp (vs. tip grasp) follows the trend $D > C > B > A$, with **D** significantly higher than **A** ($p = 0.0371$).

7.2 Subjective Metrics

NASA Task Load Index (TLX) (See Figure 11 and Table. 3)

How does the Total Score compare for conditions A, B, C and D?

Just Accepted

Fig. 11. Nasa TLX. Model fit for NASA TLX scores per expertise group.

R10: The expected **TLX metrics score** of **D** was significantly better than in conditions **A**, **B** and **C** for both DRC and ONR expertise levels (see AD, BD and CD contrasts p-values in Figure 11).

Physical demand of IC-based direct teleoperation

Just Accepted

Table 4. Significance of the model factors for the HRI Metrics

Analysis of Deviance Table (Type III Wald chisquare tests)

Pr(>Chisq)	Robot Traits	Bond	Goal	Manipulation	Perception	Satisfaction
(Intercept)	<2.20E-16 ***	<2.20E-16 ***	<2.20E-16 ***	<2.20E-16 ***	<2.20E-16 ***	<2.20E-16 ***
Condition	3.83E-07 ***	5.38E-11 ***	1.33E-05 ***	0.0007684 ***	3.09E-08 ***	0.004428 **
Task	0.943346	0.358	0.97249	0.2435082	0.005267 **	0.672051
Expertise	0.058972 .	0.1124	0.03851 *	0.02134 *	0.052819 .	0.074689 .
Condition:Task	0.89175	0.1978	0.60924	0.0233901 *	0.002505 **	0.101637
Condition:Expertise	0.004623 **	4.18E-05 ***	0.05206 .	0.0094851 **	0.888265	0.065146 .
Task:Expertise	0.75977	0.4509	0.95849	0.3074015	0.283971	0.764447

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R11: **A** and **B** resulted in significantly higher physical demand when compared to both computer-based interfaces (**C** and **D**) for both expertise groups ($p < 0.0001$).

Interaction between Condition and Expertise

R12: A number of TLX metrics had a significant 2-way interaction. **TLX Total Score** ($Pr(> Chisq) = 0.024559$): Both IC-based teleoperation conditions (**A** and **B**) had significant differences between the DRC and ONR groups (A: p=0.144; B: p=0.0267) with a better score in the ONR group, whereas the computer interface conditions (**C** and **D**) did not result in significant differences per expertise group. The expected **total score** follows the relation $D < C < A < B$ for the DRC group, and $D < A < B < C$ for the ONR group (a lower score is better). **Frustration** level had significantly better scores for the ONR group for all conditions, as well as **temporal demand**, and **performance** (see Fig.11).

HRI Metrics in Post-Condition Survey (See Figure 12 and Table 4)

R13: The expected mean score of condition **D** is better than in all other conditions and expertise levels in all the HRI metrics of the post-condition assessment, and participants rated **D** *significantly* better than the other methods for a number of the HRI scales in the different expertise groups, as detailed in the contrast p-values (AD, BD and CD) in Figure 12. The Cronbach's α of each subscale is also reported in Figure 12, with Robot Traits, Goal Subscale, Manipulation, and Satisfaction obtaining the highest values in the range 0.72 to 0.93.

Condition Ranking in Post-Experiment Survey (See Figure 13)

R14: The Cronbach's α consistency measure among all requested rankings is 0.8694. In the ONR group, the **Overall final ranking** showed no significant difference between **B** and **D** ($p = 0.4283$), both of which ranked better than **A** ($B : p = 0.0003; D : p = 0.0638$) and **C** ($B : p < 0.0001; D : p = 0.0035$). In the DRC group, **B**, **C** and **D** had no significant difference, ranking better than **A**.

8 ANALYSIS AND DISCUSSION OF RESULTS

The monotonic trade-off between task time and motion accuracy as human-robot co-activity increases can be reversed with assisted planning: The use of assisted planning in condition **D** resulted in *total task times* comparable to both IC-based direct teleoperation conditions **A** and **B** for all tasks and expertise levels (R2), whereas condition **C** resulted in significantly longer task times than all other conditions (R3). While **D** times are comparable to **A** and **B**, all other objective metrics were improved by **D** (R9), showing an objective

Just Accepted

Fig. 12. HRI Metrics. Six Likert Scales (1-7 response range) from Post-condition survey. We fit a linear mixed model to each group and report the p-values of the condition contrasts intra- and inter- expertise groups, as well as the Cronbach's α measure of consistency. Error bars show the Standard Errors.

improvement over the overall goal of the operator (to prioritize accuracy while minimizing task time). This result has implications for the previous understanding of a monotonic trade-off between execution speed and motion accuracy when moving from pure teleoperation to models of human-robot collaboration that increase the level of co-activity according to the skills of each agent. In particular, assisted planning in **D** does in fact achieve task

Just Accepted

Fig. 13. Condition Rankings from Post-Experiment Survey. Error bars show the Standard Errors.

Just Accepted

times comparable to teleoperation while improving accuracy (instead of improving accuracy at the cost of slower task times).

Increased robot autonomy removes interaction effect of task time with expertise: Even though **C** and **D** use the same computer interface, total task time in **C** had a significant 2-way interaction of task time with expertise (DRC vs. ONR) that **D** did not exhibit (R5 and R6), indicating that the use of assisted planning resulted in a performance equalizer between the two expertise groups. Task time in IC-based conditions **A** and **B** did not experience significant variations for expertise groups, possibly due to the intuitiveness afforded by the IC. This result has implications for the deployment of these systems in real high-intensity domains, where experts from different fields often come together to work on a given situation.

Task difficulty interacts with interface type: Task difficulty for each task (T1, T2 and T3), as measured by

Task Time, resulted in a different ordering for IC-based conditions (**A** and **B**) than for teleautonomy conditions (**C** and **D**). The results show that task time followed the relation $T1 < T2 < T3$ in **C** and **D**, whereas it followed $T1 < T3 < T2$ in **A** and **B** (R7). This interaction indicates that different manipulation maneuvers present different levels of difficulty in the computer interface than in teleoperation with the IC.

Total task time is heteroscedastic: The data on Total Task Time exhibited strong heteroscedasticity, which shows that the variance of the total time increases as a function of increasing time. The finding of larger variability for longer runs indicates diminishing performance predictability for longer runs, possibly due to supervening circumstances in a particular task run or the dexterity of the operator.

Task learning effect doesn't transfer across different interfaces: The main effect of Trial over task time was significant ($Trial2 < Trial1$, $p < 0.0001$), showing a learning effect for a given task on a given condition (R8). However, users did not become faster at a certain task as they repeated it in different conditions throughout the course of the experiment (manipulation check ($p = 0.34$)).

The addition of 3D perception in teleoperation resulted in longer task times, while the expected higher accuracy was not significant: The addition of 3D perception in IC-based direct teleoperation conditions resulted in a total time increase in **B** with respect to **A** (R4). The expected benefits of the 3D perception on the improvement of the other metrics related to manipulation accuracy did not result in significant changes in **A** vs. **B**. TLX scores are favorable to **A** but not significantly, possibly due to the workload associated with the management of the 3D view in **B**. Operators reported in the section for open comments that the addition of the 3D view in teleoperation was very informative. However, this addition did not result in observable benefits, possibly hindered by the complexity of adjusting the 3D view while operating the IC.

TLX metrics show a decrease in workload for assisted planning: Subjective metrics for task workload (TLX) and human-robot collaboration are favorable to **D** in all categories (R10). Computer interfaces show significantly lower physical demand than IC-based teleoperation conditions ($p < 0.0001$), also indicating that other teleoperation approaches based on body motion should be evaluated in light of this metric.

9 CONCLUSIONS

This paper presents an experimental assessment of a human-robot teaming model based on assisted planning for multi-step remote manipulation that leverages learned task models [29] to compute and suggest motion plans to the operator. We compare this system with three established models: direct teleoperation with 2D and 3D perception and teleoperation based on a 3D user interface. The study replicated real field conditions as much as possible; all aspects of the system were implemented end-to-end (perception, planning, controls, communications, user interfaces), and no Wizard of Oz technique was used. The study was conducted with an expert population in teleoperation of mobile manipulators. The following are three main results of the study: **(1)**Assisted planning achieved task times comparable with direct teleoperation through an imitation controller, while improving task time significantly over using the same interface without the assisted planning component. **(2)** Additionally, it improved a number of objective (e.g. grasp quality, collisions, regrasp), subjective metrics (NASA TLX [8] and HRI metrics [12] [9] [10]). **(3)** The use of end effector teleoperation through the 3D user interface had a significant interaction with the previous expertise of the users. The addition of assisted planning (motion suggestions from the robot), while using the same interface, removed this interaction, resulting in a performance equalizer across users.

Just Accepted

There are limitations inherent to this assisted planning approach. The discrete nature of the planning-and-execution workflow makes it suitable for quasi-static scenes and not for dynamic scenes, in which teleoperation (if done without time delay) would still allow more flexibility. Similarly, using a pre-learned task model limits the system to work with known tasks and objects, a challenge that could be further explored by using task models that can adapt on-line to variations beyond novel positions and orientations. The results of this study strongly support the advantages of **assisted planning**, validating the need for further research to increase the capabilities of the underlying models to be more flexible, generalizable, and adaptive in an intuitive and on-line fashion.

Finally, the protocol design, tasks, metrics and statistical models might serve as guidance for other benchmarking studies in teleoperation and manipulation, whether it involves human users in the loop or comparison of machine learning models for autonomous manipulation.

REFERENCES

[1] Brenna D. Argall, Sonia Chernova, Manuela Veloso, and Brett Browning. 2009. A survey of robot learning from demonstration. *Robotics and Autonomous Systems* 57, 5 (2009), 469 – 483.

[2] Mark H. Burstein, Bolt Beranek, Newman Inc, and Drew V. Mcdermott. 1996. Issues in the development of human-computer mixed-initiative planning. In *Cognitive Technology*. Elsevier, 285–303.

[3] Daniel W Carruth and Cindy L Bethel. 2017. Challenges with the integration of robotics into tactical team operations. In *Applied Machine Intelligence and Informatics (SAMI), 2017 IEEE 15th International Symposium on*. IEEE, 000027–000032.

[4] Jessie YC Chen, Ellen C Haas, and Michael J Barnes. 2007. Human performance issues and user interface design for teleoperated robots. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 37, 6 (2007), 1231–1245.

[5] Anca Dragan and Siddhartha Srinivasa. 2012. Formalizing Assistive Teleoperation. In *Robotics: Science and Systems*.

[6] Maurice Fallon, Scott Kuindersma, Sisir Karumanchi, Matthew Antone, Toby Schneider, Hongkai Dai, C. Pérez-D'Arpino, Robin Deits, Matt DiCicco, Dehann Fourie, Twan Koolen, Pat Marion, Michael Posa, Andrés Valenzuela, Kuan-Ting Yu, Julie Shah, Karl Iagnemma, Russ Tedrake, and Seth Teller. 2015. An Architecture for Online Affordance-based Perception and Whole-body Planning. *Journal of Field Robotics* 32, 2 (2015), 229–254.

[7] Philip E Gill, Walter Murray, and Michael A Saunders. 2002. SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM journal on optimization* 12, 4 (2002), 979–1006.

[8] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology* 52 (1988), 139–183.

[9] Guy Hoffman. 2013. Evaluating fluency in human-robot collaboration. In *International conference on human-robot interaction (HRI), workshop on human robot collaboration*, Vol. 381. 1–8.

[10] Guy Hoffman. 2019. Evaluating Fluency in Human–Robot Collaboration. *IEEE Transactions on Human-Machine Systems* 49, 3 (2019), 209–218.

[11] Sture Holm. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics* (1979), 65–70.

[12] Adam O Horvath and Leslie S Greenberg. 1989. Development and validation of the Working Alliance Inventory. *Journal of counseling psychology* 36, 2 (1989), 223.

[13] Siddarth Jain, Ali Farshchiansadegh, Alexander Broad, Farnaz Abdollahi, Ferdinando Mussa-Ivaldi, and Brenna Argall. 2015. Assistive robotic manipulation through shared autonomy and a Body-Machine Interface. In *Rehabilitation Robotics (ICORR), 2015 IEEE International Conference on*. 526–531.

[14] Shervin Javdani, Siddhartha Srinivasa, and J. Andrew (Drew) Bagnell. 2015. Shared Autonomy via Hindsight Optimization. In *Proceedings of Robotics: Science and Systems*. Rome, Italy.

[15] Hong Jun Jeon, Dylan Losey, and Dorsa Sadigh. 2020. Shared Autonomy with Learned Latent Actions. In *Proceedings of Robotics: Science and Systems (RSS)*.

[16] Matthew Johnson, Brandon Shrewsbury, Sylvain Bertrand, Duncan Calvert, Tingfan Wu, Daniel Duran, Douglas Stephen, Nathan Mertins, John Carff, William Rifenburgh, Jesper Smith, Chris Schmidt-Wetekam, Davide Faconti, Alex Graber-Tilton, Nicolas Eyssette, Tobias Meier, Igor Kalkov, Travis Craig, Nick Payton, Stephen McCrory, Georg Wiedebach, Brooke Layton, Peter Neuhaus, and Jerry Pratt. 2017. Team IHMC's Lessons Learned from the DARPA Robotics Challenge: Finding Data in the Rubble. *Journal of Field Robotics* 34, 2 (2017), 241–261. https://doi.org/10.1002/rob.21674

[17] Kirsten E Kaplan, Kirk A Nichols, and Allison M Okamura. 2016. Toward human-robot collaboration in surgery: performance assessment of human and robotic agents in an inclusion segmentation task. In *Robotics and Automation (ICRA), 2016 IEEE International Conference on*. IEEE, 723–729.

Just Accepted

[18] Sisir Karumanchi, Kyle Edelberg, Ian Baldwin, Jeremy Nash, Jason Reid, Charles Bergh, John Leichty, Kalind Carpenter, Matthew Shekels, Matthew Gildner, David Newill-Smith, Jason Carlton, John Koehler, Tatyana Dobreva, Matthew Frost, Paul Hebert, James Borders, Jeremy Ma, Bertrand Douillard, Paul Backes, Brett Kennedy, Brian Satzinger, Chelsea Lau, Katie Byl, Krishna Shankar, and Joel Burdick. 2017. Team RoboSimian: Semi-autonomous Mobile Manipulation at the 2015 DARPA Robotics Challenge Finals. *Journal of Field Robotics* 34, 2 (2017), 305–332. https://doi.org/10.1002/rob.21676

[19] O. Khatib, X. Yeh, G. Brantner, B. Soe, B. Kim, S. Ganguly, H. Stuart, S. Wang, M. Cutkosky, A. Edsinger, P. Mullins, M. Barham, C. R. Voolstra, K. N. Salama, M. L'Hour, and V. Creuze. 2016. Ocean One: A Robotic Avatar for Oceanic Discovery. *IEEE Robotics Automation Magazine* 23, 4, 20–29.

[20] Eric Krotkov, Douglas Hackett, Larry Jackel, Michael Perschbacher, James Pippine, Jesse Strauss, Gill Pratt, and Christopher Orlowski. 2017. The DARPA Robotics Challenge Finals: Results and Perspectives. *Journal of Field Robotics* 34, 2 (2017), 229–240. https://doi.org/10.1002/rob.21683

[21] Pat Marion, Maurice Fallon, Robin Deits, Andrés Valenzuela, C. Pérez-D'Arpino, Greg Izatt, Lucas Manuelli, Matt Antone, Hongkai Dai, Twan Koolen, John Carter, Scott Kuindersma, and Russ Tedrake. 2017. Director: A User Interface Designed for Robot Operation with Shared Autonomy. *Journal of Field Robotics* 34, 2 (2017), 262–280.

[22] Negar Mehr, Roberto Horowitz, and Anca Dragan. 2016. Inferring and Assisting with Constraints in Shared Autonomy. In *Conference on Decision and Control (CDC)*.

[23] Katharina Muelling, Arun Venkatraman, Jean-Sebastien Valois, John Downey, Jeffrey Weiss, Shervin Javdani, Martial Hebert, Andrew Schwartz, Jennifer Collinger, and Andrew Bagnell. 2015. Autonomy Infused Teleoperation with Application to BCI Manipulation. *Proceedings of Robotics: Science and Systems*.

[24] Robin R Murphy. 2004. Human-robot interaction in rescue robotics. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 34, 2 (2004), 138–153.

[25] Keiji Nagatani, Seiga Kiribayashi, Yoshito Okada, Kazuki Otake, Kazuya Yoshida, Satoshi Tadokoro, Takeshi Nishimura, Tomoaki Yoshida, Eiji Koyanagi, Mineo Fukushima, and Shinji Kawatsuma. 2013. Emergency response to the nuclear accident at the Fukushima Daiichi Nuclear Power Plants using mobile rescue robots. *Journal of Field Robotics* 30, 1 (2013), 44–63. http://dx.doi.org/10.1002/rob.21439

[26] Kirk A Nichols, Adithyavairavan Murali, Siddarth Sen, Ken Goldberg, and Allison M Okamura. 2015. Models of human-centered automation in a debridement task. In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*. IEEE, 5784–5789.

[27] Kirk A Nichols and Allison M Okamura. 2016. A framework for multilateral manipulation in surgical tasks. *IEEE Transactions on Automation Science and Engineering* 13, 1 (2016), 68–77.

[28] Adam Norton, Willard Ober, Lisa Baraniecki, Eric McCann, Jean Scholtz, David Shane, Anna Skinner, Robert Watson, and Holly Yanco. 2017. Analysis of human–robot interaction at the DARPA Robotics Challenge Finals. *The International Journal of Robotics Research* (2017), 483–513. Issue 5-7.

[29] Claudia Pérez-D'Arpino and Julie A. Shah. 2017. C-LEARN: Learning Geometric Constraints from Demonstrations for Multi-Step Manipulation in Shared Autonomy. In *IEEE ICRA 2017*.

[30] Daniel Rakita, Bilge Mutlu, Michael Gleicher, and Laura M Hiatt. 2019. Shared control based bimanual robot manipulation. *Science Robotics* 4, 30 (2019).

[31] Brennan Sellner, Frederik W Heger, Laura M Hiatt, Reid Simmons, and Sanjiv Singh. 2006. Coordinated multiagent teams and sliding autonomy for large-scale assembly. *Proc. IEEE* 94, 7 (2006), 1425–1444.

[32] Thomas B. Sheridan. 1992. *Telerobotics, Automation, and Human Supervisory Control*. MIT Press, Cambridge, MA, USA.

[33] Eliza Strickland. 2014. Fukushima's next 40 years. *IEEE Spectrum* 51, 3 (2014), 46–53.

[34] Milind Tambe, Paul Scerri, and David V Pynadath. 2002. Adjustable autonomy for the real world. *Journal of Artificial Intelligence Research* 17, 1 (2002), 171–228.

[35] Russ Tedrake. 2014. Drake: A planning, control, and analysis toolbox for nonlinear dynamical systems. (2014). http://drake.mit.edu.

[36] David Whitney, Eric Rosen, Elizabeth Phillips, George Konidaris, and Stefanie Tellex. 2020. Comparing robot grasping teleoperation across desktop and virtual reality with ROS reality. In *Robotics Research*. Springer, 335–350.

[37] G. N. Wilkinson and C. E. Rogers. 1973. Symbolic Description of Factorial Models for Analysis of Variance. *Applied Statistics* 22, 3 (1973), 392–399.

[38] Brian M. Yamauchi. 2004. PackBot: a versatile platform for military robotics. In *Proc. SPIE 5422, Unmanned Ground Vehicle Technology*, Vol. 5422. 228–237.

[39] Holly A Yanco, Adam Norton, Willard Ober, David Shane, Anna Skinner, and Jack Vice. 2015. Analysis of human-robot interaction at the DARPA robotics challenge trials. *Journal of Field Robotics* 32, 3 (2015), 420–444.

Just Accepted